

Before Each Session:

1. Make sure known author name is on first sheet, and participant name is on title of spreadsheet.
2. Fill in highlighted portions below for this participant. If multiple participants will go in a row, put them all down.
3. Put personas sheets in correct order: [1: tfidf, 2: random, 3: relevance, 4: textrank, 5: topics]
4. Insert data, noting where “more topics” columns should be.
5. Double check data in EACH sheet.
 - a. Note randomized order for tasks and methods in participant spreadsheet.
 - b. Make sure ranking methods of materials columns align with methods columns.
6. Make sure old answers are removed.
7. Insert abbreviations.
8. Just capitalize abbreviations with no replacements.
9. Hide necessary rows/columns for known author.

Are you okay with me recording this video call and transcribing it?

Part 1: Accuracy of Participant's Author Profiles (18 mins)

I will first share a link with you that presents information about an author known to you [you as an author]: [link](#)

Would you mind sharing your screen so that I can see the spreadsheet?

Repeat for topics/tasks/methods/materials {

**If Materials* but favorite methods was “more topics” column or None, skip materials section. Otherwise, unhide materials column corresponding to favorite methods column.*

Unhide necessary information/instructions.

As stated at the bottom of the sheet, Read instructions.

Initial Instructions:

- All the information you will see is based on research since 2015.
- Please answer the following questions based on your own knowledge of this author.

Topics Instructions:

- On the left is a Topics column describing the author's research.
- Please highlight the topics that are unclear, are too generic, or do not reflect the author's research well.

Tasks Instructions:

- On the right are 5 potential columns to further describe the author's research. The column headers describe the kind of information provided in each column.

- Please rank the 5 columns in terms of how well they **complement** the first column to more accurately reflect the author's research.
- Please take into account the terms themselves, how they are ordered, and how they work together to give a full representation of the author's research.
- Label the best column with a 1, the second best with a 2, etc.
- If none of the columns complement the first column to more accurately reflect the author's research, do not rank the columns and type "None" below.
- For your highest-ranked column, please highlight the terms that are unclear, are too generic, do not reflect the author's research well.

Methods Instructions:

- On the right are 5 potential columns to further describe the author's research. The column headers describe the kind of information provided in each column.
- Please rank the 5 columns in terms of how well they **complement** the columns already shown to accurately reflect the author's research.
- Please take into account the terms themselves, how they are ordered, and how they work together to give a full representation of the author's research.
- Label the best column with a 1, the second best with a 2, etc.
- If none of the columns complement the initial columns to more accurately reflect the author's research, do not rank the columns and type "None" below.

Materials Instructions:

- On the right is a Materials column to further describe the author's research.
- Does the Materials column **complement** the columns already shown to accurately reflect the author's research? Below the Materials column, type "Y" if yes and "N" if not.
- Please take into account the terms themselves, how they are ordered, and how they work together to give a full representation of the author's research.

[**If Tasks** Feel free to reread those instructions and] let me know when you're ready to start.

Please think aloud about how you are ranking the author profiles [deciding to answer the question **if Topics/Materials**].

Set timer for 2 mins if Topics, 8 minutes if Tasks, 5 minutes Methods, 3 mins if Materials.

The layout for their task ranking methods is: tfidf, topics, random, relevance, textrank

**if Tasks* The profile they ranked highest for tasks round: relevance*

if not Topics I would like to ask you a few more questions about the column you chose [information in this column **if Materials**].

- Are there any terms that stand out as particularly helpful/unhelpful in describing the author? Why?
- Are there any important terms to describe the author that are missing from the column?
- What do you think about the granularity of the terms in the column? Are they too broad, too narrow, or just right?
- What are your thoughts on the order of the terms in the column?

Additional questions to potentially ask:

- What do you think are important factors to consider when ranking your top Tasks/Methods?

}

Part 2: Usefulness of Unknown Authors' Profiles (19 mins)

The 5 next spreadsheets will be used to present 5 different unknown authors. Please go to the spreadsheet labeled "Author 1."

This is an unknown author with whom you are not supposed to be familiar.

Repeat for each of 5 unknown authors {

As stated at the bottom of the sheet, Read instructions.

Tasks Instructions:

- Does the Tasks column **complement** the Topics column to help you understand what kind of research the author does? Below the Tasks column, type "Y" if yes and "N" if not.
- Please take into account the terms themselves, how they are ordered, and how they work together to give a full representation of the author's research.

Methods Instructions:

- Does the Methods column **complement** the Topics and Tasks columns to help you understand what kind of research the author does? Below the Methods column, type "Y" if yes and "N" if not.
- Please take into account the terms themselves, how they are ordered, and how they work together to give a full representation of the author's research.

Materials Instructions:

- Does the Materials column **complement** the Topics, Tasks, and Methods columns to help you understand what kind of research the author does? Below the Materials column, type "Y" if yes and "N" if not.
- Please take into account the terms themselves, how they are ordered, and how they work together to give a full representation of the author's research.

if 1st author Does that make sense?

Please think aloud about how you are deciding to answer the questions.

Set timer for 3 mins.

Please go to the next sheet in the spreadsheet.

}

when done rating the unknown author profiles I would like to ask you a few more questions about the unknown authors' profiles. Feel free to move back and forth among the 5 unknown author spreadsheets.

Set timer for 7 mins.

- Are there any terms that you don't know? If so, what was your reaction to seeing these terms?
- Would you have liked to see terms that are more related to your own interests ranked higher, even if they were not central to the author in question?
- What do you think about the granularity of the terms in these profiles? Are they too broad, too narrow, or just right?
- Are there any terms that stand out as particularly helpful/unhelpful in understanding what kind of research the authors do? Why?

Additional questions to potentially ask:

- Is this information helpful for deciding whether or not to learn more about an author? Why or why not?
- Is there any other particular context in which you think this information would be helpful or interesting to you? (e.g., deciding if should chat with them, deciding if they are a potential collaborator, etc.)
- Is there any other information not shown that you would like to see about the author?

Part 3: Accuracy of Participant's Personas (5 mins)

Please go to the spreadsheet labeled Personas [1: tfidf, 2: random, 3: relevance, 4: textrank, 5: topics].

In this final part of the experiment, we show you clusters automatically extracted by finding different groups of authors who tend to collaborate with Author X. We take each group and their papers and show you the main topics/tasks in that group. This can help tease apart the different things X works on with different people.

Since these authors can overlap across groups (think of how you can belong to multiple groups), you may see the same topic [task] appearing in multiple groups, too.

The groups are not meant to be totally distinct — overlap is expected.

More senior researchers tend to engage in different kinds of research over the course of their careers. In other words, they can wear different hats that correspond to different lines of research. Here, we represent their different hats as distinct personas.

As stated at the bottom of the sheet, *Read instructions*.

Instructions:

- This sheet presents the author's overall profile on the left as well as their different personas on the right. The personas are meant to represent the author's different lines of research.
- Please rate your agreement with the statement below from 1-5. 1 means you strongly disagree, 5 means you strongly agree.
- Please take into account the terms themselves, how they are ordered, and how they are distributed among the personas to give a full representation of the author's research.
- STATEMENT: "The personas reflect the author's different research interests (since the year 2015) well."

Feel free to reread those instructions and let me know when you're ready to start.

Please think aloud about how you are deciding to answer the questions.

Set timer for 5 mins.

when done answering the questions, as time permits I would like to ask you a few more questions about the personas.

Additional questions to potentially ask:

- How would you describe each of these personas in your own words?
- How distinct are these personas? Do they overlap in any way?
- Would you say one persona is the author's dominant persona? Would you say some personas are more important than others?
- What do you think of the number of personas presented? Should the author have more or fewer personas?
- What do you think about the granularity of these personas? Are they too broad, too narrow, or just right?
- What do you think about the ordering of the personas?

That is everything. Thank you very much for participating.

You should receive your compensation within 2 weeks. If you don't, please let me know.